A Reference Model for Information Management to Support Coalition Information Sharing Needs

Mark Linderman (POC)
Air Force Research Laboratory
Information Directorate
525 Brooks Road Rome, NY 13441-4505
mark.linderman@rl.af.mil

Barry Siegel
SPAWAR Systems Center - San Diego, Code 2461
53560 Hull St. San Diego, CA 92152-5000
barry.siegel@navy.mil

David Ouellet
Defence Research and Development Canada - Valcartier
2459 Pie-XI Blvd North
Val-Belair, Quebec, Canada G3J 1X5
david.ouellet@drdc-rddc.gc.ca

Julie Brichacek
Air Force Research Laboratory
Information Directorate
525 Brooks Rd Rome, NY 13441-4505
julie.brichacek@rl.af.mil

Sue Haines
DSTL Information Management
Malvern Technology Centre
St Andrews Road
Malvern Worcs WR14 3PS UK
slhaines@dstl.gov.uk

Greg Chase
Command & Control Division
Defence Science and Technology Organisation
PO Box 1500 Edinburgh
South Australia 5111
greg.chase@dsto.defence.gov.au

Janet O'May
U.S. Army Research Laboratory
ATTN: AMSRD-ARL-CI-CT
Aberdeen Proving Ground, MD 21005-5067
janet.omay@us.army.mil

ABSTRACT

Effective information management is a key enabler in ensuring that commanders have access to the information they need to make optimal decisions. This can only be achieved by the combination of people, processes and tools to form an effective system. Many systems and technologies claim to address the information management problem. However, in the absence of a universally accepted model of information management, it is hard to assess how well these products and systems support military information management requirements. Simply stating that information management is the capability to provide the right information, in the right place, and at the right time, does not provide a yard stick against which to measure a system's capability. The Technical Cooperation Program (TTCP) Command, Control, Communications and Information (C3I) Action Group (AG) 4 has taken on the challenge of defining a model of information management to support command and control (C2) in a coalition context. The proposed model provides an abstract representation of the important elements that are required to form an effective information management system. This has been validated by the examination of tools that support information management and the model's coverage of their capabilities.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2005		**00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE **A Reference Model for Information Management to Support Coalition Information Sharing Needs**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory,Information Directorate,525 Brooks Road,Rome,NY,13441-4505**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**19**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

1. INTRODUCTION

The goal of Information Management is to maximize the effectiveness of an enterprise (e.g., military operations) by maximizing its ability to act upon information that is produced and consumed within the enterprise and externally. There are several means by which this can be accomplished:

- Reducing barriers to effective information use by providing notification, mediation, access control, and persistence services
- Providing an information space wherein information is managed directly, rather than delegating all information management responsibilities to applications that produce and consume information
- Focusing on consumer needs rather than producer preferences to ensure that information can be effectively used
- Providing tools to assess information quality and suitability of information
- Exploit producer-provided characterization of information to support automated management of information.

There are many definitions of information management extant, many of which apply generic management terms to the information domain. An example is OMB Circular A-130[1] that defines information management as "the planning, budgeting, manipulating, and controlling of information throughout its life cycle." Our TTCP[2] action group's interest is in defining a model of information management that is technical in nature, and therefore we will not focus on the planning and budgeting concerns, but rather on manipulation and control of information and other facets that are significant. Others[3] refer to the "treatment" or "skillful handling" of information. While not inaccurate, these terms are too vague for the purposes of defining a model of information management that is technical in nature. We define information management as *a set of intentional activities to maximize the value of information to support the objectives of the enterprise.* These activities include the promulgation of standards, control of information and the administrative activities that support it, and the capture, dissemination, manipulation, persistence and destruction of information. The goal of this paper is to describe these activities and the actors that perform them or are affected by them.

There are several aspects of this definition that bear upon the structure of the information management model (discussed in Section 2). The first is the word "activities." Managing is not an end state or a frame of mind, it is something one does. Therefore, the model enumerates and describes these activities that comprise the *verbs* of management. Implicit in the definition are actors that interact with the managed information environment by either managing information or by sharing managed information. Finally, there is the information itself that, together with the actors, comprise the *nouns* of information management.

The next word that bears scrutiny is "intentional." Arguably, one might consider unintentional (de facto) activities, but we emphasize intention precisely because there appears to be a prevalent attitude that information management is "everywhere and nowhere." Our premise is that enterprises are not maximizing the value of the information they possess because either 1) information management is viewed as "someone else's" responsibility, or 2) it is delegated down to individuals or organizations without sufficient authority or resources to perform it effectively.

Information management is a "set" of intentional activities that act in concert with one another to achieve a desired result. If these activities operate independently of one another, the results may be minimally acceptable but far short of optimal. For example, access control and information transformation may act together to sanitize information in accordance with disclosure policies prior to sharing it with a given consumer. Acting independently, the access control activity may be forced to decide between passing all or none of an object. Therefore, passing nothing might adhere to

[1] http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html

[2] "TTCP is an international organization that collaborates in defence scientific and technical information exchange; program harmonization and alignment; and shared research activities for the five nations, Australia, Canada, New Zealand, the United Kingdom, and the United States. The Command, Control, Communications and Information Systems (C3I) Group is responsible for co-operative and collaborative development of technologies to achieve interoperable, seamless information systems focused on the support of allied military missions." Under the C3I Group, the charter of the Information Management action group AG-4 is to examine the area of information management at the strategic, operational and tactical levels within a coalition force.

[3] www.impact21group.com/glossary.htm, lorien.ncl.ac.uk/ming/cleantech/glossary.htm

the prevailing security policy, but operational effectiveness may suffer because finer-grained enforcement is not an option.

The definition explicitly states that information role is to support the enterprise. Information seldom has a value in itself; rather, it is a tool to achieve other things. Therefore, the *value* of information is not an intrinsic quality, but dependent upon how it is used. To effectively perform value-based management, the managers and perhaps the infrastructure must be able to estimate the value of information. To do this, it is important to understand the business processes supported by that information and the relative importance of those processes. Ultimately, it is desirable to prioritize resource allocation at a high level, perhaps at the process level, and have the priorities of lower-level objects (e.g., supporting information) be automatically derived from it. This is an example of management granularity that has a direct impact upon the effort required to manage.

The information management model presented (Figure 1) is an abstract representation of the essential activities of information management and the actors that interact within an information environment to achieve their objectives. The purpose of the model is to lift the quality of dialog about information management from generalizations and vaguely stated capabilities to thoughtful consideration of specific activities and appropriate assignment of resources, roles and responsibilities. Perhaps most importantly, we hope to convey that information management should be an enterprise-level challenge; one that is too important to ignore or relegate to users that are insufficiently resourced, lacking authority, or lacking appreciation of enterprise-level objectives.

We recognize that any information management model needs to include the following critical elements: people and their roles; processes; and implementation technologies. Our model is equally applicable to a range of 'systems', from manual information exchange processes to wholly automated systems. An information management system, which implements this model can be constructed from a range of technologies and the model itself should not be seen as a constraining process that mandates a particular architectural style e.g., publish and subscribe architectures, or service oriented architectures. Rather it should be regarded as identifying the information management requirements that would need to be supported by these architectures in order to provide an effective information management capability.

In a coalition context, there are several implications of information management that we assume to hold. These assumptions are:

1. Acquisition cannot be standardized or synchronized. Perhaps a stronger statement of this is that acquisition should not be standardized or synchronized, if the result of doing so is inability to interact with new coalition partners, unacceptable delays in fielding capabilities, or unacceptably high costs for fielding or supporting capabilities.
2. Universal data standards are unrealistic. While progress can certainly be made on broad agreements of both high level data standards (e.g., standard upper-order ontologies) and within domains (e.g., Command and Control Information Exchange Data Model (C2IEDM)), insistence upon a single, universal model is counterproductive.
3. Data standards version skew is the norm not the exception. While this assumption is primarily a by-product of asynchronous acquisition, it is worth explicitly noting because it has specific implications on the information management environment for coalition operations. If we assume that different participants will migrate to newer standards at different rates, then it is incumbent upon information managers to deal with the problem.
4. Coalition Information Space must adapt quickly to new partners and processes. Similar to assumption 3, this assumption affects where adaptation is likely to take place. Given current software technology, especially as seen in complex military applications, it is unrealistic to assume that applications will be able to adapt to changing processes, products or partners. Today adaptation is carried out by changing the behaviour of people; often through patchwork fixes and workarounds. From an information management perspective, this situation can be ameliorated by providing managed information tailoring environments and tools to support processes that can be modified by policy rather than with software that cannot.

In short, we must learn to accommodate diversity in mission, motivation, technical infrastructure, and allegiance and culture. To a large extent, an information management environment, which forms a bridge between the participants in a coalition operation, provides an ideal place to accommodate diversity. It provides one place in an information architecture where changes should be made. With this in mind, it is essential that our information

management environments be designed with adaptability in mind.

2. A MODEL OF INFORMATION MANAGEMENT

The model describes actors that interact with one another through the information space, the information management infrastructures that comprise an information space, and the layers of services that perform the activities of information management. Actors interact with the information management infrastructure by producing and consuming information or by managing it. External information spaces that interact with a given space are also actors from the perspective of the space. The information space is a collection of information catalogues and repositories that form the heart of an information management infrastructure. A service within a layer performs a specific information management activity. An artifact is a piece of information (e.g., a managed information object) that is acted upon by a service or that influences the behaviour of the service (e.g., a policy). We begin by describing the salient features of managed information and then describe the three components that comprise the model: actors, service layers and the information space. Another model[4] of information management served as a departure point for the model we present. Our model differs in that it

- Makes are clear distinction between actors and the services they employ;
- Addresses federates to support diverse communities (such as coalitions);
- Organizes activities according to type of service (i.e., service layer) rather than by process steps (e.g., task, submit, consume)
- Is more explicit on maintenance and security activities, and significantly enhances transformation services; and
- Supports flexible workflow rather than imply a specific information process through which all information must pass from production to consumption.

2.1. Managed Information Objects

Informal words, such as "document" and "object", are often used to describe a quantum of information. In a sense, the managers of information and information management infrastructures only want to know enough about a quantum of information to do their jobs; they do not need to know everything about it. They need a characterization of the information. For example, if a movie clip shows a car chase, the managers do not need to have a detailed description of each frame; it may be enough to know that the topic is a car chase, and that it occurred at a given time and location. Furthermore, given the volume of information a typical enterprise uses, it is important that the characterization be readily available.

Because we are interested in the management of information, we need to define more formally what these objects are. The quantum of managed information is called a managed information object (MIO). A MIO comprises a payload (e.g., the movie clip in the example above) and metadata that characterizes the object (e.g., topic, time, and location). It is desirable that all of the information needed for making management decisions (such as content-based dissemination) be present in the metadata in a form that permits efficient processing. While there may be cases where this is not possible (e.g., payload inspection for virus detection), the degree to which it holds may have a direct bearing upon the efficiency of management. An important element of characterization is the concept of type. The type of an object (e.g., "satellite imagery") is useful for determining how the information should be characterized and for setting policy on its appropriate use.

2.2. Actors

The information management model depicts a set of actors who interact with the information infrastructure and are known as consumers, producers, federated spaces (collectively referred to as "federates"), and managers. Depending on their role, these actors will consume or produce information, or will manage the infrastructure. A description of the mechanisms and products that each actor will either generate or require in order for information management system to operate effectively follows.

2.2.1. Consumers

Consumers are information space clients that utilize the information within the information space. A consumer requests information from the information space by presenting their security credentials and transaction request. Within a transaction request, the consumer describes details such as the type of information, search criteria, subscription criteria, prioritization of returned results, format for information to be returned in, and length of subscription. Consumer's transaction requests are typically filled by the information space from either

[4] Information Management Capability Development Roadmap (IMCDR), Mike Farrington, March 2004 - QINETIQ/KI/IntSys/PUB040917

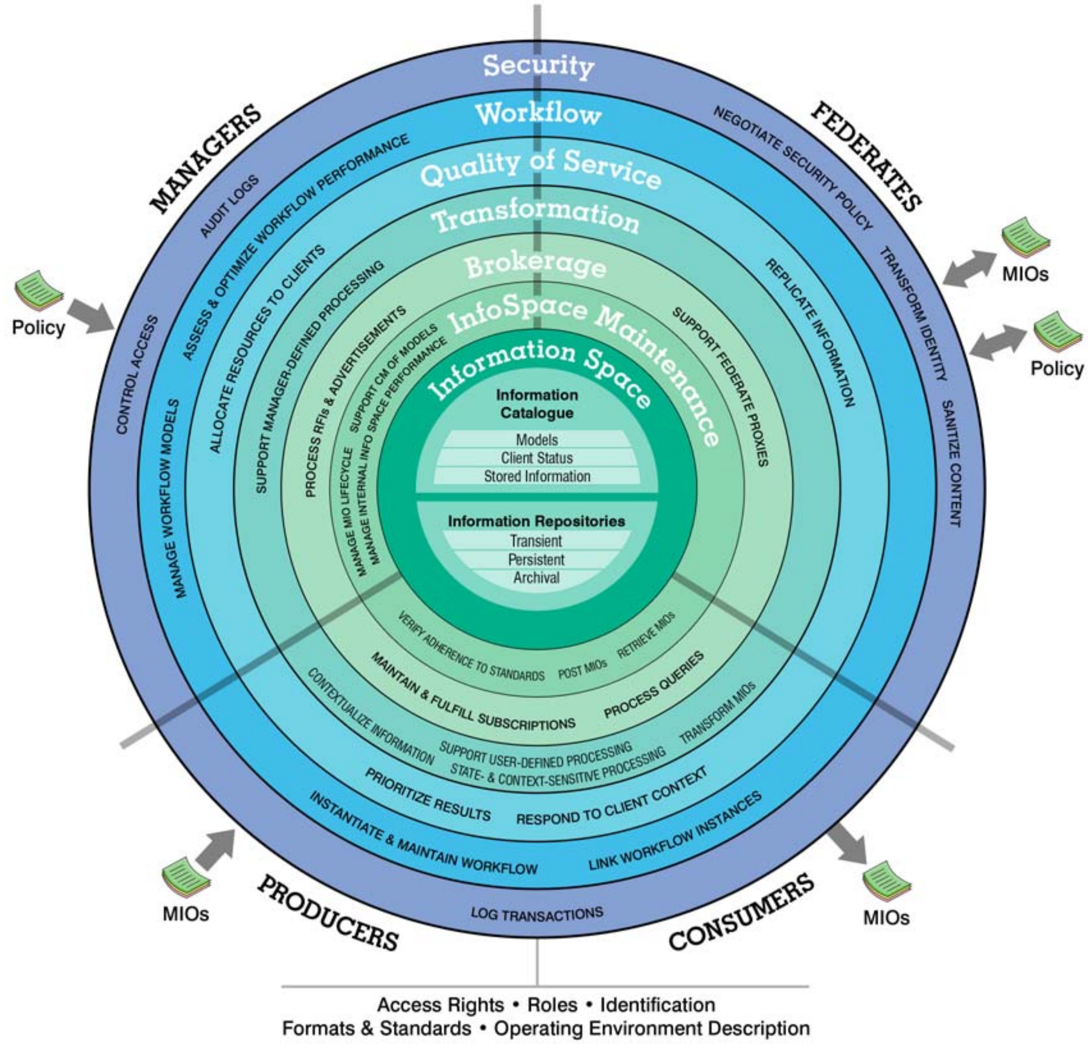

Figure 1. Information Management Model

the information catalogue and/or information repositories. The types of transactions a consumer may want are the ability to browse, query, subscribe, transform and assess suitability of information within the information space.

Search/Browse: Consumers browse the information space if they want to see what is currently contained in the information space. Browsing includes the ability to search and receive information from the information space and is supported by the brokering layer. The brokering layer should support several modes of search, including simple keyword searches similar to Google™. A search returns to the consumer links to available MIOs that meet the search criteria. Search and browse are less structured, both in the predicate language and the process, then subscription and query (defined next).

Subscribe: Consumers want to be notified of new or updated information as it becomes available. Subscription requests form a part of the Information Catalogue. A request states what information is required and for how long the subscription should be fulfilled by the brokering layer. The brokering layer should support the capability of accepting, setting and fulfilling subscriptions. Once the brokering layer accepts a subscription, notification of MIOs that meet subscription's criteria should be sent as they become available.

Query: Consumers query the information space to retrieve previously published information. A query is more structured than searching in that the search criteria are expressed in predicates based upon the structure (and potentially semantics) of MIO metadata. The brokerage layer fulfils queries by retrieving either a partial or full set of MIOs.

Transform: Consumers will want the ability to filter information, set prioritization of received results and have information reformatted to their specific operating environment. These types of user-defined manipulations are set by users, sent to the information space as part of browse, query or subscribe and then performed on the resultant information by the transformation layer. An information space that supports user-defined manipulation of information will allow for processing of the information retrieved from the space prior to presentation to the consumer.

Suitability, Assessment, and Feedback: Consumers should perform suitability assessments and provide feedback on the information that has been received from the information space. Today, assessment is typically a manual process carried out by users, based upon experience, that rates information in terms of its fitness for the user's needs. In some instances some elements of the assessment process can be delegated to the consumer process acting on the user's behalf, for example, where initial assessment can be made based upon previous user preferences or ratings. Consumer feedback should be provided to the information space so that managers can collect, sanitize, and then send feedback to the specified producers. A consumer may submit a Request For Information (RFI) if its requirements are not being met.

2.2.2. Producers

Producers are information space clients that add information to the information space. The information will reside in the persistent or transient information repositories, depending on the requirements. In most cases the producer client will have created this information via some other process, for example a remote sensing system. In some cases a consumer client may recover information from the repository, add value via some algorithm, and then add processed information back to the repository. Such a client is both a producer and a consumer. The functions of a producer are to publish information, advertise capabilities, process feedback from consumers, receive requests for information, and remove incorrect information.

Publish: A producer's primary interaction with an information space is publication of information. For published information to be effectively managed, the producer needs to characterize the information published, using terms that can be interpreted by both the managers of the information space and consumers. At a minimum, this characterization should include the time of publication and the identity of the publisher. For products of a particularly sensitive nature, where the provider is unwilling to delegate access control decisions to the information space, the provider may opt to publish only the metadata together with a pointer to the information directly held by the provider. This method of publication into the space may also be used for 'oversized' information items or access to real-time data feeds.

Advertise: Producers may describe products that it could (if requested) produce. For example a given producer may have the capability to obtain an aerial image of a given place. The advertisement of the capability should include a description of the available product, possibly constructed from a collection of metadata or through the use of an agreed ontology. This advertisement should also provide the client with information that can enable it to invoke any processes, external to the information space, that are required for the producer to provide such capability.

Receive Consumer Feedback: Producers may be able to offer a better service to consumers if they receive some feedback from the information space. This feedback should, at a minimum, provide retrieval statistics for the information produced by the producer, and could provide qualitative feedback on the relevance of the information supplied to the consumers needs. For example, consumers might indicate that the granularity of the information supplied should be improved, suggest an alternative format for the information, highlight the need for supply of information relating to a different geographic location, or request that information be packaged into smaller chunks.

Receive Targeted Requests for Information: Producers may receive targeted requests for information, based upon RFIs (which may be related to the advertised capabilities of the information producer). The receipt of an RFI may trigger the collection and subsequent publication of relevant, characterized information. RFIs may also be used by the information producer to guide their future production priorities.

Retraction of Published Information: A producer may need to retract information that is later deemed to be incorrect. This retraction should not remove the information from the repository (as it was actually published and available to decision makers at some point in time) but instead mark the information as "retracted".

2.2.3. Federates

The ideal of a single, universal information space is unlikely to be achieved in any sufficiently complex

military or commercial endeavour, such as in coalition operations. Wherever it is necessary to have multiple information spaces, it is likely that clients will want to reach across them to get the information they need. There are two non-mutually exclusive methods of interaction that may be employed to accomplish this: a client may simultaneously connect to each information space directly or the spaces themselves may be connected within a federation that allows one information space to be 'extended' by the contents of another.

The second model of interaction, that of having the spaces interact directly (rather than through client interfaces) requires higher-level service level agreements between the respective managers, but it has the potential for interactions that are more seamless and better managed. In this model, a producer need not make a special effort to share information with other information spaces, it merely shares it with its 'native' information space and this space ensures that it is shared appropriately with other spaces. The advantage is fairly clear: federation services connect information spaces together for the purpose of providing consumers a "single" information space. However, the challenges to establishing the necessary interfaces and agreements bear further discussion. The federation interfaces to the information space encompass trust management, confidentiality and integrity management, policy mediation, content filtering, information replication and pass-through processing.

Coalition federates exacerbate concerns that, while not unique to a coalition environment, do warrant discussion. Among these concerns are trust and security and access to information catalogues.

- **Trust and Security:** While coalition partners may be assumed to be trusted, coalition systems are often not. Therefore, a coalition environment will need to enforce national-specific multiple security levels. Federate interfaces should support traditional approaches for discretionary and mandatory access, provisions for security downgrades, content filtering, and on-the-fly auditing.

- **Controlled Catalogue Access:** Coalition information management is typically asymmetric; one country may choose to disseminate a piece of information that another would not. Even the mere fact that a country holds a particular piece of information may not be releasable to another. This means that not only must MIOs be subject to access control, but also the fine grained information contained in catalogues.

Federates are outside information spaces and, by extension, so are their managers and the clients they serve. In this regard, federates are like producers and consumers, except that as 'outsiders,' there are likely to be more restrictions on interactions with them. As 'external' actors, they do not have control over the information space. Examples of activities that federates need are seamless information access, explicit restrictions on usage (bi-directional) of information, mediation of inconsistent data standards, ensure shared information integrity, and information service-level agreements.

Seamless Information Access: Subject to necessary constraints, federates want seamless access to information within the information space. This means that they want their consumers to get information from within the information space without additional effort. They want information that they share with the information space to be shared with its consumers subject to restrictions set forth below. Essentially, while there may be many policies that have to be negotiated between the managers of federated spaces, in the end, clients should not be aware that a seam exists.

There are several mechanisms that may be employed to do this: one is a complete replication of information from one space to another; a second is periodic requests for information; and a third is establishing 'standing' information requests (similar to consumer subscriptions) that result in sharing a subset of the available information. The third option implies an ability to accept subscriptions from other information spaces. An extension to the third option is to allow consumer query requests to be "passed through" to the federates for processing.

Explicit Restrictions on Usage (Bi-directional): An information space is likely to be trusted by the producers and consumers that use it, but it may not be trusted by another information space. Even if there is trust in the space itself, a federate may impose usage restrictions on information that it shares. These restrictions may limit who can access the information and the ability of the recipient to pass the information on to third-party information spaces.

Mediation of Inconsistent Data Standards: Different information spaces are more likely to have different data standards and interoperability requirements. Therefore, a federate would like to have information transformed to serve the needs of its community. Regardless of whether this transformation takes place before or after transmission from one space to the next, it requires, minimally, access to the syntax and semantics in use in the other federate. This implies that there must be

a mechanism to request this information. Additionally, it implies that within the transformation layer of one or both of the federates that the transformation be performed, according to rules defined by managers of the spaces.

Ensure Shared Information Integrity: A federate may need mechanisms to ensure that recipients of its information can determine that information it is providing has not been altered. Furthermore, as clients within a federate make reference to a MIO from a federate, they need some mechanism for consumers of the MIO to dereference it.

Information Service-Level Agreements: Much of the foregoing discussion has focused upon what a federate may do to an information space, but not what an information space must do. These obligations are often formalized by contractual information service-level agreements (SLAs). An SLA commits a federate to a stipulated level of service. An SLA may specify requirements like regular updates to certain types of information. Therefore, an SLA could specify trust and confidentiality requirements, quality of service expectations in terms of timeliness of response to requests and the degree of relevance of information, mandatory feedback from the federate at predefined intervals, and requirements for information recall. For example, an SLA could stipulate that not only shall the receiving federate be notified of an MIO recall, but also all other federates be notified who have received related or derived MIOs.

2.2.4. Managers

Managers are responsible for monitoring and controlling the information. Management services are those set of processes that allow the information space to be managed by a staff. These services enable the information space to process information demands in a timely, reliable manner through judicious use of policies, resource management and maintenance. A key element for management services is ensuring performance by managing the lifecycle of the individual MIOs. To use the information space within a coalition environment, it is vital that access control policies be set and enforced. Manager responsibilities are as follows: navigate and understand the syntactic relationships among MIO types; dictate and understand prevailing security policies on MIO types; monitor and control resource allocation and performance; ensure accurate data mediation; configure and monitor effectiveness of information support; establish and maintain federated space relationships; maintain information space currency; and audit information infrastructure transactions.

Navigate and Understand the Syntactic Relationships among MIO Types: The information space is a collection of information that is categorized according to type. Therefore, for managers to understand the space, they must be able to navigate the collection of types and understand how they support the goals of the enterprise. They must determine the extent to which relationships between types should be used to define and propagate policies.

To achieve this, the managers must maintain the information catalogue of the information space. Specifically, the managers must be involved in the introduction of new types. When clients create new information object types, they must describe these objects to the information space to ensure proper handling of the new MIO type; part of the description may include relationships to other types. Based upon the description provided, managers must decide whether to permit the new object type, and determine what policies will apply to it, perhaps based in part on how the MIO type relates to other types.

Additionally, managers may need to assign ancillary information to MIO types to be stored within the information catalogue. For example, this may include syntax checkers, payload renderers, and contact information for the owners of the MIO type standards.

Dictate and Understand Prevailing Security Policies on MIO Types: Managers are responsible for setting security policies for the information space. With respect to MIOs, this policy is likely to be defined at the level of the MIO type. As mentioned above, policy may be explicit and unique to a type, or inferred based upon the type's relation to other types. Generally security policy ultimately relates a user to the actions that he or she may perform on information of a specified type. There are many ways in which this may be specified, such as role-based access control described below in the Security Layer section. In a 'static' role-based system, policies define privileges that are granted to roles and then users are mapped to one or more roles. These mappings are the responsibility of managers. In a 'dynamic' role-based system, these mappings may change as a function of time or operational context.

To ensure that the maximum utility of information is realized, managers must balance security concerns against operational imperatives. Balancing constraints is fundamentally the role of a manager. However, the degree of expressiveness in defining and enforcing policy bear direct relation to the subtlety with which the manager may balance competing interests. Within the domain of security

policies, context- and content-based policies access control policies are examples of increased expressiveness that may allow more precise control of information to maximize utility without violating security constraints.

Monitor and Control Resource Allocation and Performance: Managers are responsible for ensuring that available resources are used effectively. In an information environment these resources may include distributed storage services, networking infrastructure, and computational assets. A manager must be able to monitor (at least indirectly) a system in order to control it. Therefore, key components of the system must either report to or support status queries from managers. Additionally, the managers must have some understanding of the value of the information being managed to ensure that resources are meeting users' needs. Often, this is best achieved by feedback from the users.

Ensure Accurate Data Mediation: Managers must understand how information standards apply and evolve over time. Specifically, the way in which communities of interest use information changes over time, and the syntax and semantics of the information they share changes to reflect that. However, it is often unrealistic to assume that all users will switch from older versions to newer versions in lockstep. Even if they did, there may still be information in the repository of the older format.

To ensure that information is not sent to clients in a format with a version that exceeds the latest they can understand, it is necessary for newer information to be transformed into older formats for 'legacy' systems that need it. For example, managers may classify type revisions as either 'major' or 'minor': minor revisions being those that can be converted into an older version within the same major revision. Major revisions are those that are too different from older version and cannot be converted. Note that these differences may be either syntactic or semantic in nature.

Configure and Monitor Effectiveness of Information Support: Information is seldom an end in itself; generally it is a means to achieve higher-level objectives of an enterprise. Managers need to understand the extent to which the information they manage is accomplishing these objectives. In a well-run enterprise, processes exist not only to directly support its objectives, but to improve the means by which information managers support them. One mechanism is feedback from producers and consumers of information. Managers can use this information to influence the behaviour of these actors or to modify the policies and resource allocations within the information management infrastructure.

Establish and Maintain Federated Space Relationships: As information spaces are federated together, managers are responsible for negotiating the terms of participation in the federation. This includes defining policies on access control, usage restrictions, data sanitization, and identity anonymity.

Maintain Information Space Currency: Managers may be responsible for ensuring that the information space does not become cluttered with irrelevant or outdated information. They are responsible for maintaining the performance of the information space that might become impaired if too much information accumulates in the space. Whatever the reason, the managers must manage information space repositories and the movement of information through them. Specifically, managers determine if information is put into the persistent and archive repositories and when it is removed from them.

Audit Information Infrastructure Transactions: Managers should promulgate policy that information sufficient for auditing transactions with the information infrastructure be logged for *post facto* analysis. These transaction include not only client interactions (e.g., MIO publication), but management interactions as well (e.g., modification to access control policies). Managers must have access to these logs to establish culpability in the wake of unfortunate events and to measure metrics of performance over time.

2.3. Service Layers

The activities of information management are organized into service layers. While the activities are essential to the model, the layers are solely intended to add clarity to the model; other organization schemes can certainly be developed. The activities within the layers, each a verb phrase, is an action that is performed. Generally an activity will interact with actors, the information space, and/or other activities. The ordering of the layers, while notional, does generally indicate a reasonable ordering of activities that may occur as MIOs flow from producers into the space and then back out to consumers. The information management model depicts six service layers: security, workflow, quality of service, transformation, brokerage, and information space maintenance. Each service has a specific set of activities that enable the information management system to effectively manage and provide value-added processing to information within information space.

2.3.1. Security Layer

Published products will include security and release metadata. It is the responsibility of the producer of the information to associate appropriate security and release conditions, which must be enforced by the information space. Release conditions should be specified in terms of nationality and role.

Publication of information to the space must not degrade the confidentiality, integrity and availability of the published information or the service provider publishing the information. In particular, the space must be able to preserve information integrity by ensuring that publishers authenticate to the space (so that the origin of information is not 'spoofed') and that information, once placed in the space, cannot be tampered with either by the producer or a third party. In addition, the space needs to ensure that clients are appropriately authenticated so that there is an appropriate level of confidence that information is only released in the manner prescribed by the release conditions.

The space should also have defence mechanisms to prevent hostile clients from launching denial of service attacks against information providers through its infrastructure. However, the main focus of the security layer is upon the provision of security services within the information domain, not at the networking layer. The primary functions of the security layer are controlling access, logging transactions, auditing logs, negotiating security policy with federates, transforming identity, and sanitizing content.

Control Access: The security service layer is responsible for ensuring that only appropriately authenticated producers, consumers, managers and federates interact with the information space. In addition, individual actors will be allowed to perform differing sets of interactions dependent upon their role and the management specified access control policy. In a dynamic coalition, this policy may vary over time. This reflects the expansion or contraction of the coalition and the evolution of the individual roles over the life of an operation.

Access control policy sets procedures and rules for those interacting with the information space, i.e., consumers, producers, managers and federates. The access control policy needs to be able to set and revoke access rights based on current roles of participants.

Information space participants must be assigned a recognized role, or set of roles, prior to interaction with the space. These roles should be based upon externally recognizable characteristics, such as nationality, specific job function, or location – either nationally determined or as specified in the coalition operating procedures. Additionally, workflow models may identify or refine particular roles in particular coalition contexts. A role will not usually be unique to a user, thus role based access control is coarser grained than user based access control.

Roles, as defined above are used as the atomic unit to which access rights are assigned. Access rights are specified in terms of allowed transactions with the information space for particular roles. Examples of transactions include publishing information to the information space, browsing or subscribing to the information space, or creating new information object types. An allowed transaction may be specified in terms of the type of information that a particular role may retrieve from the space; or alternatively types of information that a provider may publish to the space.

Log Transactions: Logging is the process of collecting significant management, federate, producer and consumer actions in accordance with management policy to support subsequent audit and performance tuning. Logging can include collection of who accessed the information space, for what purpose, what information was provided or recovered and when the transaction took place.

Logging policy is set by management and specifies the level of detail that must be logged for individual transactions with the space. The policy should also specify the longevity of the logging information together with an appropriate archival policy.

Audit Logs: This is a management activity, probably specified by policy but not necessarily automated, that investigates the trail of activities captured by the information space's logs. These logs can be audited for security reasons, after action review to determine which actors had access to what information at a particular time, and also in support of workflow process audits.

Negotiate Security Policy with Federates: It is the responsibility of management, supported by the security layer, to negotiate appropriate access and information replication between federated information spaces. The result of the negotiation process will be codified in federation security polices, which may require client identity to be removed before onward transmission of requests; the removal/modification of MIOs prior to replication of information to the federate space; or the modification of MIOs in response to queries from federated spaces. All of these actions should be logged and be subject to audit.

Transform Identity: The security service layer has a role in protecting the identity of clients, when their requests are passed to a federated space and also in

passing feedback to producers. In both cases, the actual role played by the client and their specific identity are not required either by the federated space which will respond to the request or by the producer receiving feedback on their product. This type of function can be considered as similar to name address translation at the network layer, i.e., the end point does not know the identity of the starting point of the connection, but the communication is still bi-directional.

Sanitize Content: The security service layer also plays a role in the sanitization of content. Here sanitization is taken to mean the modification of MIOs, in line with security policy, prior to transmission to a less trusted federated space or consumer.

2.3.2. Workflow Layer

Enterprise workflows are the patterns of information flow that support the processes of the enterprise. The steps in a workflow may be manual or automated, but typically the execution of a step in a workflow involves consuming information from previous steps, incorporation of new information and production of new information. Depending upon how steps are initiated and terminated, workflows may be categorized as either ‘event-driven’ or ‘schedule-driven’ or a combination of the two. Generally, the execution of a workflow proceeds according to a set of predefined rules with users and applications participating according to their roles. Workflow instances may have a lifetime ranging from minutes to days depending on the complexity and the duration of the various constituent activities.

There is an important distinction between workflow models and workflow instantiations. Workflow model definition occurs at “design-time,” while workflow enactment occurs at “run-time.” Workflows may be altered on-the-fly during “execution-time” so they are dynamic and subject to change. Workflows may also be “ad-hoc,” in which case they are designed and executed on-the-fly during run-time. Like other lifecycle artifacts, workflow models should be managed and placed under access and configuration controls. Transitions between states during a workflow execution should be subject to logging and auditing.

Finally, workflows may be hierarchically decomposed into sub-workflows that execute concurrently or sequentially. If a MIO is shared between workflows, then the MIO is visible to them concurrently.

Incorporation of workflows within the model is not to suggest that information management system must be workflow “centered” but rather it may be workflow “enabled” or workflow “sensitive.” The primary functions that enable effective enterprise workflows: managing workflow configurations; instantiating and maintaining workflows; linking workflow instances; and assessing and optimizing workflow enactment performance.

Manage Workflow Model Configurations: Workflow model configuration management is a set of processes to define, maintain, alter, and archive workflow models. These models are stored in the information catalogue. Archived models support *post-facto* analysis for later analysis after enactment for the purpose of workflow evolution and improvement.

Instantiate and Maintain Workflows: Previously defined workflow models are instantiated and enacted during run-time. During enactment, users are assigned roles by the security layer based on the selected workflow. As a workflow executes, its state, and those of its sub-workflows is maintained, monitored and audited.

One of the primary advantages of using workflow is the realization of the potential for workflow improvement. The ability to audit, instrument, and track a workflow provides input to refining the workflow process. A workflow stopped and started at any executable step. Stopping a workflow blocks it from further execution until it is restarted. Note that the security manager is responsible to allocate privileges allowing control of executing workflows. A generic reference model for workflow was published by the Workflow Management Coalition (1995)[5] and later adapted by the Object Management Group (2001)[6]. Our information management model has adapted key aspects of these models.

There are issues of workflow monitoring in a coalition environment. Countries may be reluctant to have their work measured, monitored, or controlled by a workflow. Therefore, workflow should rarely be mandatory. Moreover, coalition partners will likely be using different workflow tools, platforms and processes. For these reasons, workflow interoperability may be a consideration during design of a workflow model.

---

[5] "The Workflow Reference Model", Document Number TC001-1003 by the Workflow Management Coalition, 19-Jan-95 Author David Hollingsworth found at http://www.wfmc.org

[6] Workflow management Facility Specification, V1.2" Object Management Group (OMG), April 2000, found at http://www.omg.org, OMG Domain Specifications

Link Workflow Instances: Executing workflows can be linked on-the-fly during run-time to create new workflows. Workflow instances may be either sub-workflows or related workflows. MIOs may be shared between sub-workflows but may not be shared between related workflows.

Assess and Optimize Workflow Performance: Workflow processes are rarely defined perfectly the first time. Experiences and metrics derived from previously enacted workflows can be used for workflow redefinition and improvement. This activity encompasses the assessment of previously enacted workflows for the purposes of workflow model improvement.

2.3.3. Quality of Service (QoS) Layer

The Quality of Service (QoS) layer is responsible for ensuring that an information management system can best serve the needs of its consumers, producers and federates, when faced with changing operating conditions. A task made more difficult as these clients can't always clearly specify what QoS attributes they require from the information management environment.

Quality of service can occur at different levels and it is important to differentiate the QoS at the information level from the QoS of the underlying communication mechanism. While it is desirable to implement an information management system over a communication system that has a high level of QoS, having poor QoS in the communication domain does not preclude having a high QoS in the information domain. Having good QoS in the information management system would mean that when the communication system manages to send a piece of information, it is more likely to satisfy the needs of the client. Conversely, having low QoS in the information domain but high QoS of the communication layer would mean that the client would reliably receive information that might not meet its needs. The activities that can be performed in an information space to increase QoS in the information domain are: respond to client context, allocate resources to clients, prioritize results, and replicate information.

Respond to Client Context: In order to provide a context for their QoS requirements information clients may need to be able to specify their environment. This may include performance metrics like:

- How many objects per second could they create?
- How many objects per second are they prepared to consume?
- How many objects per second do they need
- What bandwidth do they expect to receive through the information space?
- What bandwidth from the information space can they deal with?

Allocate Resources to Clients: Resources need to be managed to ensure that consumers receive an appropriate level of service from the information space both in terms of timeliness of response to requests and subscriptions; and also in terms of quantity of information. Similarly, producers need to be able to interact with the space in an efficient manner. This means that the QoS layer should ensure that feedback from the space is delivered in a timely manner and that the producer is not overloaded by RFIs. The QoS layer needs to take into consideration consumer constraints such as low bandwidth links or end-user device limitations.

The QoS layer is also responsible for ensuring that producers can submit information to the space as a non-blocking activity. This means that the producers' availability should not be adversely affected by interactions with the space. Similarly, a consumer submitting feedback information to the space should be able to submit this without impacting on performance.

Prioritize Results: The information space must be able to support consumer preferences for ordering of information. This ordering may be defined by some form of user defined manipulation function that should reside within the transformation layer. In addition, information prioritization should aim to maximize the amount of 'real' information passed to the consumer, e.g., if two adjacent MIOs between them match a consumer's geographically related query and a third MIO, which matches the query, overlaps with the previous MIOs and adds no further new information, then only the first two MIOs should be sent.

From the perspective of an information producer, the information space would be expected to prioritize RFIs being sent to producers. This prioritization should take into account any consumer supplied priority information, policy set by managers of the space, the ability of the producer to respond to the RFI both in terms of bandwidth constraints and the amount of outstanding concurrent requests for information.

Replicate Information: The QoS layer is responsible for ensuring that appropriate MIOs are replicated to federated information spaces. This function is required to meet service level agreements between federated information spaces, provide resilience in the information domain by providing

multiple repositories for critical information, and to ease consumer pressures on particular information spaces for frequently requested information. The information replication process needs to be managed so that replication is scheduled to take into account of the need minimize the impact on the performance of the information space whilst still ensuring that information is replicated in a timely manner.

2.3.4. Transformation Layer

In an information management environment, the role of the transformation layer is to adapt information as it enters or leaves the information space in order to satisfy client needs and management policies. The primary activities within the transformation layer are contextualizing information, transforming MIOs, state- and context-sensitive information processing, and managing and executing user-defined information processing requests.

Contextualize Information: Information context comprises the manner in which information relates to other information, workflows, and enterprise objectives. While most consumers may be content to consume MIOs in their native format, others may desire information to be set in context prior to delivery. This may require that elements of this context be combined with the MIO. Typically, the user may define what context is of interest and how it is to be represented; this is an example of a user-defined processing function. Other alternatives include manager-defined contextualization or clients that perform this as a value-added service.

An additional function of contextualization is to evaluate newly posted information. This facilitates the work done at the brokering layer such as providing clients with information closely related to their needs by supporting prioritization decisions and determining suitability to address specific needs.

Transform MIO: Even though there may be a one-to-one correspondence between published MIOs and information delivered to a consumer, the consumer may not want to consume the information in the same form as it was produced. When requested by a consumer, MIO may be modified in order for that piece of information to fulfil the needs of the consumer and to meet the constraints specified by the managers of the information space. Therefore, the transformation layer may transform MIOs into more readily transmitted and consumed formats. This may entail the transformation of the payload, the metadata, or both. These transformations can take several forms. For instance, some of the transformations that could be performed on satellite imagery are:

- Changing format to send it in a format usable by a client software
- Cropping the image to only cover a specific region
- Lowering the resolution to meet security related policies

Support State- and Context-Sensitive Processing: Context-sensitive processing functions are activated depending on the operating context of either a single client or the information space as a whole. The role of these functions is, to a certain extent, to automatically configure the information space based on external inputs and pre-established conditions. One example of a processing function affecting a single client is to lower the resolution of images sent to it when the client's connection to the information space operates in a degraded condition. A more wide spread impact might occur if a new coalition partner or NGO is allowed access to the space, which would affect management policies including acceptable format for release of different types of MIOs.

State-dependent processing depends not only upon a processes current input but also upon its past history. The information space operates in a dynamic environment, so it is necessary for processing functions to maintain their status while waiting for new information to process in order to prevent redundant accesses to the information repositories, duplication of computation and superfluous transmissions to consumers.

Support User-defined Processing Functions: The information space must allow users to define their own processing functions to adapt (tailor) the information space to their specific needs. Users that want information transformed (e.g., manipulated, translated, aggregated, or combined) should be able to define a transformation function and have the computational and communication resources for its execution managed by the information space. Whether the resultant information is available to only that client or to a broader group is a management decision.

There are several reasons to support this activity: to reduce tedious manual manipulation, information overload, and network load. The first reason, of course, reinforces the second. The third recognizes that not only are consumers burdened by 'non-decision quality' information but so is the network.

Support Manager-defined Processing Functions: The information space is a dynamic environment and new types of information and format get added. The purpose of manager-defined information processing functions is to maintain transformations allowing the

mediation of one MIO version to another and reconcile inconsistencies between the formats. These functions are activated when sending information to a user that requested a different version of the MIO. Manager defined information processing functions can also be invoked to satisfy security restrictions by the modification of MIOs so that the information released is consistent with security policy.

2.3.5. Brokerage Layer

The role of the brokerage layer is to match available (or potentially available) information to information needs. Brokering of available information involves activities that reflect the means by which consumers seek information: processing queries, notifying actors (principally consumers), and maintaining and fulfilling subscriptions. Note that the brokering layer does not have a significant role in search and browse interactions because in the mode, consumers are largely 'self-brokering' with the assistance of Google-like indexing that is supported by the information catalogue. Brokering for information that is not currently available is more complicated as it involves processing consumers' requests for information and producers' advertisement of potential products. Finally, to support federates that desire seamless interaction with the information space, the brokerage layer may support federate proxies that handle consumer requests forwarded from other information spaces. The brokerage layer is responsible for processing queries, maintaining and fulfilling subscriptions, processing requests for information and supporting federate proxies.

Process Queries: A query request is a well-structured request that retrieves information for a consumer based upon a predicate that is evaluated against the metadata of MIOs in the persistent information repository. There are three appropriate responses to a valid query: notification that no available information meets the criteria of the query, a complete set (either references or entire MIOs) matching the request, or a partial result set. The first two 'complete' the transaction, but partial results are more complex because consumers should be able to 'page' through partial results sets until they are satisfied. In any case, once the brokerage has a result set, it may pass it to the transformation layer to tailor and prioritize it prior to passing it back to the consumer.

Maintain and Fulfill Subscriptions: Subscriptions are on-going requests for information that are satisfied as new information becomes available. Consumers (i.e., subscribers) subscribe to information by describing what they need, generally in terms of the type of information and predicates over the MIO metadata to filter out irrelevant information. These subscriptions are stored as part of the client status information catalogue. As a new MIO is posted by a producer to the information space maintenance layer, the brokerage layer is notified so that it can determine which subscribers are interested in it. The subscribers can then be notified. To meet the tailoring requirements of a subscription, it may be necessary to pass MIOs through the transformation layer prior to sending them on to the subscriber.

Process Request for Information (RFI) and advertisements: Finding ways to satisfy unmet demands for information require awareness of the demands, potentially available information products and the means to influence the production of that information. The client status information catalogue at the core of the information space maintains information about both consumer needs and producers' potential products. Based upon this information, RFI processing may identify potential producers, and forward RFIs to them, sanitized if necessary. Ideally, producers will acknowledge if they will meet the needs and if it appears that a need will go unsatisfied, then managers may be brought in to rectify the situation. The managers must determine if a need is valid, prioritize it for resource allocation and then, using their awareness of how to influence the production information, attempt to make producers responsive to it.

Support Federate Proxies: A federate information space may choose to delegate the fulfilment of one of its consumer requests (whether it be a query, RFI, or subscription) to the information space so that information may be shared seamlessly. While simple in concept, the complex trust relationships that may exist between the spaces (and the communities they serve) complicate things. To handle these complicated issues, rather than having federates interact directly with brokerage activities, the brokerage layer supports federate 'proxies' that do so, in a controlled manner, on their behalf. These proxies may perform tasks such as masking the identity of the requesting consumer or aggregating subscriptions to reduce duplicate transmissions to the federate. It may also have to sanitize queries, subscriptions, RFIs and advertisements prior to passing this information on to a federate.

2.3.6. Information Space Maintenance Layer

The information space maintenance layer interacts directly with the underlying information space. It is responsible for: posting new information to the information space repositories; verifying the format (e.g., adherence to standards) for posted information; informing the brokerage layer of the introduction of

new MIOs; managing the lifecycle of the MIOs in the repositories – including retraction of MIOs by their producers; managing internal information space performance; and providing support for configuration management of models stored in the information space (e.g., work flow models and MIO types). Finally, it is responsible for retrieving specific MIOs from the information space repositories.

Post MIOs: When an information producer posts a MIO to the information space, the information maintenance layer is responsible for the insertion of the MIO into the correct information repositories. If subscription for new information is supported for a specific type, the broker layer must be notified of the existence of new information. It may then determine which consumers are interested in the information. Once identified, the information may be queued for delivery from the transient repository to the consumers.

Verify Adherence to Standards: It is desirable that information posted to the information space adhere to prevailing standards. These standards may include naming conventions, proper syntax, encoding, et cetera. In an efficient information management environment, verification will be an automated process based upon rules promulgated by the information managers. For example, the metadata for a MIO may be encoded in extensible markup language (XML) and validated against a XML schema appropriate to that type of information. Furthermore, the payload may be validated for adherence to standards.

To verify adherence with standards, information managers must first register both the standards and the mechanisms for verifying compliance with them. Managers store this information in the information models catalogue where it is associated with the appropriate MIO types. When MIOs are posted to the information space by producers or federates, their adherence to prevailing standards can then be verified before the MIOs are placed into the repositories.

Manage MIO Lifecycle: Within the information space, MIOs are stored in one of three places: persistent repository, transient repository, or in an archive. The transient repository holds MIOs to support subscription servicing. The persistent repository holds MIOs to respond to queries. Finally, the archival repository is for long-term off-line storage. It is the responsibility of managers to define policies for the movement of MIOs between these three repositories. Once a MIO has been removed from the transient and persistent repositories, it is no longer available to the brokerage layer. Management action is required to retrieve MIOs from the archive.

MIOs may be retracted by authorized managers or clients. Many implementations will destroy the retracted information while others will tag the retracted information accordingly. Prevailing policies related to auditing may require that posted information not be destroyed or altered. We generally believe that it is not a good idea to destructively update information and that it is preferable to post a new MIO and record the fact that it supersedes an earlier MIO in the metadata of the new object or externally in a third MIO, or by tagging the superseded MIO.

Manage Internal Information Space Performance: As an internal activity, the information maintenance layer manages the internal information space performance. There are many means of doing this, the details of which are certainly implementation-specific. One means of doing this is to constrain the size of the information repositories through appropriate archiving to limit the growth of the 'on-line' transient and persistent repositories. Other options include physical placement of information in a distributed system, and configuration of clustering caching mechanisms and partitioning of system resources by MIO type. Since the 'internal' organization of the information space bears upon the quality of service of the system, there is likely to be interaction between this activity and the QoS layer.

Retrieving specific MIOs from repositories: As many MIOs are likely to be transmitted by reference (e.g., a hyperlink), it is necessary that there be a means of retrieving a MIO from a repository based upon such a reference. Similar to web servers today, there are a number of potential ways to request information and several exceptional conditions that may result. Unlike a traditional web page, MIOs have both metadata and (optionally) payload, and consumers (or services) may desire one, the other, or both.

2.4. The Information Space

Within the information domain, the information space comprises the set of managed information, regardless of where is it physically resides. However, the placement of the information, who has control over it, and the operations they can perform have a significant impact upon the effectiveness of management and the value gleaned from the information. In other words, just because information exists within the space that does not mean that those who need it and who should be authorized to get will necessarily get it. Nor does it mean that those who have information to share will be able to effectively do so. The suboptimal utilization of information often arises because

stewards of information do not feel empowered or motivated to share that information, or simply because consumers do not know how to ask for it.

While the model does not distinguish between implementations that make management activities easy or hard (or impossible), it does suggest that a coherent management strategy is important. Furthermore, while not represented in the model, we assert that technologies are emerging that allow an information space to be constructed in ways that either facilitate or allow automation of many activities of information management. In particular, we should be able to build information spaces that are distinct from producer's and consumer's applications. Such an approach enables management of information by those able to do so in a coherence manner matched to enterprise objectives. This is seldom achieved by end-user applications. The 'coherence' information space is not necessarily centralized, nor is it oblivious to the needs and constraints of information producers and consumers. It does, however, enable the emphasis to be on information and the demands for that information rather than working around the limitations of information producers.

The information space comprises information repositories and an information catalogue that enables consumers' and producers' information requirements to be met. The activities within the service layers within the model enable the information space to be managed and interact with producers, consumers and other spaces.

2.4.1. Information Catalogue

The role of the Information Catalogue is to maintain all the meta-information related to the use of the information space and the information contained therein. The following describe the different types of meta-information needed to adequately manage information.

Stored Information: The stored information catalogue contains the index of all the accessible information in the local repositories. This catalogue is used when querying the information space. Another use for this catalogue is to allow published information objects to reference one another. Reiser's law of information economics states that the expressive power of an information system is proportional not to the number of objects that get implemented for it, but instead to the number of possible effective interactions between objects in it[7]. In the case of an information management system, this expressiveness is related to the number of possible references between information objects.

An important kind of metadata for information objects that needs to be expressed is related to detailing the provenance of information – the information pedigree. Information pedigree is necessary for mediating policies in a federation of information spaces, for instance, when a new piece of information is derived from diverse information elements that have different release conditions. Information pedigree is also necessary for consumers to determine if a piece of information corroborates another one or if they originate from the same source.

Models: The Information Models catalogue contains the taxonomy of the different MIOs supported by the information space. This catalogue can be organised hierarchically to support specialisation of MIOs. This catalogue allows consumers to discover information categories that are of interest and to browse, subscribe to, or query for them. The Information Models catalogue also serves as the basis for the description of user-defined transformation as well as for the tailoring of the information from producer specified information models to consumers specified ones.

Workflow models are also stored within the models catalogue. A workflow model is a template for interaction that allows certain patterns of activity to be coordinated more easily. Instances of these templates can be specialised for specific needs.

Client Status: The Client Status catalogue contains information that describes the current status of the producers and consumers that interact with the information space. While client status may contain information not directly related to the management of information (e.g., unit operational readiness), we focus upon status information that is. Specifically it may describe products that producers are capable of producing and information that consumers desire.

Consumer status may contain descriptions of its needs, particularly those that are not adequately met by the information space. It may also specify constraints (such as available bandwidth) that affect its ability to consume information. Producers can query this catalogue to find potential consumers for products they can produce. This information is necessary to prevent producers from having to publish volumes of information, even though there may be no demand for most of it, simply because there is no mechanism by which they can determine consumers' needs. The request for information could contain the type of information needed, using the Information Models catalogue, as well as other

[7] ReiserFS – http://www.namesys.com/v4/v4.html

metadata such as the area of interest and the time period.

Producer status information is complementary to the consumer status information. It allows producers to describe the information products they can make available. The taxonomy of possible products needs to be described in the Information Models catalogue to allow the Information Space to match consumers' needs with providers' capabilities. Also, a description of information products permits consumers to select a preferred producer for a particular type of information. This selection of producers can also be based on policies stating authoritative sources for specific types of information.

Another use for the Information Providers catalogue is to include the list of Subject Matter Experts (SMEs) on different topics. These topics can range from a specific type of information described in the Information Models catalogue to higher level domains such as amphibious operations planning. The information provided by SMEs might not transit through the Information Space but the fact that this information was requested should.

2.4.2. Information Repositories

The role of the Information Repositories is to receive, maintain and provide the actual pieces of information or MIOs. The different types of repositories are transient, persistent, and archival. The repositories roughly correspond to the lifecycle of the MIOs they persist. The transient information repository is a temporary holding place for newly published information. The persistent information repository holds 'active' information so that it can be queried for. Finally, the archival information repository holds 'inactive' information that is no longer directly available for querying by clients. The policies describing the transition of information object from one repository to another are set through the information maintenance service of the management interfaces.

Transient Information: Transient information is information from a producer that is in the process of being relayed onto consumers; it is 'passing-though' the information space. Consumers must subscribe to transient information in order to receive it. An information space may use a transient repository to hold transient information until it has been delivered to all consumers. Transient information is not retained on a long-term basis by the information space and in particular consumer processes cannot query against the transient repository. An information space may copy some or all transient information into a persistent or archival repository according to information maintenance policies. In some systems, such as peer-to-peer systems, the transient information may be stored on the producers, web caches, or intermediary nodes in a virtual overlay network.

Persistent Information: Persistent information is information from producers that is retained by the information space as defined by the management policy. Unlike transient information that serves consumers that have expressed interest for information prior to its production (via subscription), persistent information serves those that wish to search/browse, or query for it after the fact. The persistent information repository is the working repository of the Information Space. To avoid having the size of the repository grow to the point where performance (or information quality) suffer unacceptably, it is likely that information will have to be removed from it after a time. Ideally, this is subject to policy based upon how the information is used within the enterprise.

Archival Information: Information that is in the archival repository is not directly accessible to consumer processes. A separate management process must be invoked to recover archived material. An information space may have a policy that defines when information is moved from the transient or persistent repository repositories to the archival repository. Typically, this off-line storage is used for auditing purposes.

3. HOW TO APPLY THE MODEL

Many systems and technologies claim to address the information management problem. However, in the absence of a detailed model of information management, it is hard to express military information management requirements and to assess how well products and systems support requirements. Simply stating that information management is the capability to provide the right information, in the right place, at the right time does not provide a yard stick against which to measure a systems capability (it is about as useful as specifying that a system be 'user friendly'). Indeed, the real questions are who is going to perform what information management activities, with what resources, and will this support the operational objectives of the enterprise.

The model, as defined in section 2.0, provides an abstract representation of the important actors and activities that are required to form a functional information management system. There are several ways to apply the model, such as assessing existing technology/systems, specifying requirements for new systems, and understanding capability gaps. The

remainder of this section describes two applications of this model: assessing existing technology/systems and specifying requirements for new systems.

3.1. Assessing Existing Technology and Systems

The model can be used to assess the degree to which a specific set of technologies support information management requirements. Generally we are interested in the flexibility of these systems, the degree to which they automate information management activities, and the ease and expressiveness of actors' interactions with the technologies.

The model has been used to assess two technologies: the Microsoft® SharePoint™[8] and the Air Force Research Laboratory Joint Battlespace Infosphere[9,10] (JBI) reference implementation[11]. SharePoint[12] is a content management portal. The JBI is an information management capability envisioned to support military needs. The fact that one of these is a commercial tool while the other is a science and technology (S&T) project highlights the fact that an information management model can be used in different phases of the acquisition process. A chief information officer may need to evaluate SharePoint versus other commercially available content management systems. The research laboratory or commercial company may look to the model to find which elements of the model maximize the return on investment of R&D resources.

Without providing the detailed analysis here, it is clear that the different systems emphasize different parts of the model. Both systems support Publish, Subscribe, Query and Browse. SharePoint excels at management of more static, page-oriented information. Users can be alerted to updates to individual documents, libraries of documents, or lists, but consumers primarily rely upon search and browse interactions. These strengths reflect typical uses of information portals. JBI, in contrast, primarily focuses on dynamic information; primary consumer interactions are subscription and query.

[8] SharePoint™ is a trademark of Microsoft Corporation.

[9] United States Air Force Scientific Advisory Board, "Report on Building the Joint Battlespace Infosphere, Vols. 1 and 2," SAB-TR-99-02, December 17, 1999, available from http://www.sab.hq.af.mil/Archives/index.htm

[10] http://www.rl.af.mil/programs/jbi/

[11] Combs V., et. al., "Joint Battlespace Infosphere: Information Management within a C2 Enterprise," to appear in ICCRTS 2005, CCRP.

[12] http://www.microsoft.com/sharepoint/

JBI is built around the information space models catalogue that holds schema (and other related information) about the set of MIO types supported in the system. This allows JBI to have type-based, in addition to role-based, access control. SharePoint does not implement a model catalogue and therefore is not type based. Its access control is based upon role and location in the site collection. JBI's type based system relies more on queries over structured metadata whereas SharePoint relies more upon full text indexing augmented by document metadata.

Perhaps less obvious are the underlying assumptions about lifetime of information. JBI places a premium on rapid posting and delivery of new information with new objects being published at potentially sub-second rates. SharePoint works at time scales more amenable to human processes. In the model, the difference is evidenced by JBI's emphasis on the transient repository. Both JBI and SharePoint require the persistent repository.

JBI is developing a sophisticated implementation of the transformation layer activities of the model[13], whereas SharePoint has better support for incorporating pre-existing information sets through linking to external file systems.

Despite these differences, there are several similarities, most notably in sections of the model that they do not support well. Neither system has addressed the heterogeneity- and security-related aspects of federation well. Neither system organically supports producer/consumer feedback mechanisms. Both have only limited QoS management capabilities to date. SharePoint has a limited workflow capability built in, third party tools are much better. JBI requires third party solutions for workflow. Likewise, both systems require external tools to create and manage the archival repository. More significantly, both currently lack automated means of removing information from the persistent repository as the value of the information decreases with time.

If the model is being used to assess available commercial tools such as SharePoint to support acquisition decisions, decision makers must decide if lacking capabilities are important, and if so whether they will be supported through third-party tools that require integration or whether they will remain manual activities that incur recurring training and manpower resources. If the model is applied to a technology development effort such as JBI, it can help prioritize investment; for example emphasizing a new browse activity over automated feedback to producers.

[13] http://www.fuselet.org/tech-overview/

3.2. Specifying Requirements for New Systems

Management invariably involves striking balances between competing concerns. In an enterprise that relies upon information management – almost all enterprises do, whether they explicitly acknowledge this or not – some activities are more important than others, and when they compete for resources, choices must be made. These resources include lifecycle costs, infrastructure (e.g., networking) limitations, and staffing requirements.

The proposed information management model provides a technology neutral description. Thus it is equally applicable to a range of 'systems', from manual information exchange processes to wholly automated systems. An information management system, which implements this model can be constructed from a range of technologies and the model itself should not be seen as a constraining process that mandates a particular architectural style e.g., publish and subscribe architectures, or service oriented architectures. Rather it should be regarded as identifying the information management requirements that would need to be supported by these architectures in order to provide an effective information management capability. These requirements allow one to specify, rank and weight each activity.

Relating requirements to technology capabilities allows one to identify, and perhaps quantify, capability gaps. Technologies may be scored for 'information management maturity,' similar to how NASA's technology readiness levels (TRLs) measure technology maturity. For example, the most primitive information management environments must support posting and browsing of unstructured information. A slightly higher level may require controlled access and publish/subscribe of typed but unstructured information. A third level may require querying and metadata-characterized MIOs. These three levels provide basic access services. The next tier adds capabilities to control performance with policy such as quality of service and format/granularity mediation within the transformation layers. Finally, the highest information maturity levels support continual process refinement and optimization. This encompasses both the Workflow layers and the feedback activities of information management.

Another application of the model in a military context is in planning of joint/coalition exercises/experimentation as it provides a framework for identifying pieces of capability that could be brought together to provide an information management capability in a coalition environment. In a sense this is little different from any other acquisition effort, but these exercises are often very budget-constrained, feature come-as-you-are system heterogeneity, and require that nations and communities of interest come together quickly to accomplish specific goals. The lack of a common information management definition, let alone lexicon and defined set of activities, makes it difficult to address information management requirements coherently. This model may serve as a basis for planning "intentional activities" that need to be supported to achieve the objectives of the exercise or experiment.

4. CONCLUSION

The proposed model provides an abstract representation of the important elements required to form an effective information management system. The model captures the principal activities needed to maximize the value of information to support the objectives of an enterprise. The model has been validated by the examination of a commercial content-management tool (Microsoft SharePoint) and a developmental prototype (the JBI Reference Implementation) that support information management and the model's coverage of their capabilities.

The proposed model provides a technology-neutral description of information management; it is equally applicable to a range of systems, from manual information exchange processes to wholly automated systems. An information management system that implements this model can be constructed from a range of technologies, and the model itself should not be seen as a constraining process that mandates a particular architectural style (e.g., publish and subscribe architectures, or service oriented architectures). Rather, the model should be regarded as identifying the information management requirements that would need to be supported by these architectures in order to provide an effective information management capability.

The model also has applicability in the planning of joint/coalition exercises/experimentation as it provides a framework for identifying pieces of capability that could be brought together to provide joint/coalition information management. A natural extension of this work is to define a set of 'information management maturity' levels that describe the efficiency of an information management implementation. Further work is needed to define a universally recognised set of levels and to apply these in a coalition context.